Федеральное государственное бюджетное образовательное учреждение высшего образования

«Московский государственный университет имени М.В. Ломоносова»

Факультет биоинженерии и биоинформатики

УТВЕРЖДАЮ
И.о. декана факультета биоинженерии
и биоинформатики МГУ
_____ / А.А. Замятнин /
«___» _____ 20__ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

«Компьютерные технологии в поиске новых лекарственных средств»

Уровень высшего образования:

магистратура

Направление подготовки:

33.04.01 «Промышленная фармация»

Направленность (профиль) образовательной программы:

«Фармацевтическая разработка»

Форма обучения:

очная

Рассмотрено и одобрено Учёным советом факультета
биоинженерии и биоинформатики МГУ
(протокол № 12 от «12» 11 2024 г.)

Москва 2024

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования (ОПОП ВО) – программ магистратуры по направлению подготовки 33.04.01 «Промышленная фармация», утвержденным приказом МГУ от 8 сентября 2023 г. № 1180.

Год (годы) приема на обучение: 2025, 2026.

1. Место дисциплины (модуля) в структуре ОПОП ВО

Дисциплина (модуль) «Компьютерные технологии в поиске новых лекарственных средств» относится к вариативной части ОПОП ВО (фиксированная дисциплина профиля), раздел учебного плана – Блок 1 «Дисциплины (модули)». Дисциплина реализуется на 1, 2-м курсе в 2 и 3-м семестрах.

Цель освоения дисциплины: формирование знаний и практических навыков применения методов биоинформатики, хемоинформатики, молекулярного моделирования и машинного обучения для поиска и оптимизации новых лекарственных средств.

Задачи дисциплины:

– освоить работу с базами данных белков, лигандов и биологической активности;

– освоить методы структурной биоинформатики: моделирование структуры белков, поиск сайтов связывания, молекулярный докинг, молекулярную динамику;

– освоить методы хемоинформатики: дескрипторы, сходство, QSAR, виртуальный скрининг, прогнозирование ADMET;

– сформировать навыки применения машинного обучения и генеративных моделей в дизайне молекул.

2. Входные требования для освоения дисциплины (модуля), предварительные условия

Базовые навыки работы в ОС Linux и программирования на Python (рекомендуется); знания органической химии и биохимии.

3. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников

Специализированные компетенции МПК приведены по реестру магистерской программы «Фармацевтическая разработка». Действия по дисциплине раскрывают вклад дисциплины в освоение компетенции; отдельные коды индикаторов МПК не вводятся.

Компетенции	Индикаторы достижения компетенций	Планируемые результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций
ПК-1. Способен определять цель и задачи исследования, подбирать соответствующую методологию исследования, составлять план исследования, демонстрировать системное понимание области исследования и предлагать методы решения поставленных задач по фармацевтической разработке	ИД-ПК-1.2. Подбирает методологию и составляет план исследования ИД-ПК-1.3. Предлагает и обосновывает методы решения задач фармацевтической разработки	**Знает:** методы компьютерного поиска и дизайна биологически активных соединений **Умеет:** выбирать и применять вычислительные методы для решения задачи поиска АФС **Владеет:** навыками проведения виртуального скрининга и анализа результатов

Компетенции	Индикаторы достижения компетенций	Планируемые результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций
ОПК-3. Способен организовывать и проводить научные исследования в области обращения лекарственных средств	ИД-ОПК-3.2. Применяет широкий набор информационно-поисковых систем для поиска научной информации	**Знает:** информационные ресурсы структурной биологии и хемоинформатики **Умеет:** осуществлять поиск и извлечение данных из PDB, UniProt, ChEMBL, PubChem **Владеет:** навыками работы с базами данных
ОПК-4. Способен к анализу, систематизации и представлению данных научных исследований в области обращения лекарственных средств	ИД-ОПК-4.2. Выбирает и применяет методы обработки, оценки и представления результатов научного исследования	**Знает:** методы оценки качества моделей (валидация QSAR, метрики ML) **Умеет:** интерпретировать и представлять результаты моделирования **Владеет:** навыками визуализации и представления результатов
ПК-3. Способен получать новые научные и прикладные результаты и представлять результаты профессиональной деятельности в устной и письменной форме в соответствии с нормами и правилами, принятыми в профессиональном сообществе	ИД-ПК-3.1. Получает и интерпретирует новые научные и прикладные результаты ИД-ПК-3.2. Представляет результаты профессиональной деятельности в устной и письменной форме в соответствии с нормами профессионального сообщества	**Знает:** принципы воспроизводимости вычислительных исследований **Умеет:** оформлять результаты вычислительного исследования в виде отчета **Владеет:** навыками подготовки научного отчета
МПК-4. Умение применять полученные знания и навыки для научных исследований в сфере выявления новых мишеней и перспективных кандидатов в лекарственные препараты	Оцениваемое действие по дисциплине: связывать вычислительный отбор кандидатов с планом экспериментальной проверки.	**Знает:** методы и ограничения, необходимые для решения комплексной задачи по дисциплине **Умеет:** связывать вычислительный отбор кандидатов с планом экспериментальной проверки **Владеет:** навыками обоснования решения, документирования результатов и оценки ограничений

4. Объем дисциплины (модуля)

Объем дисциплины (модуля) составляет 4 з.е. (144 академических часов), в том числе 120 академических часов, отведенных на контактную работу обучающихся с преподавателем, и 24 академических часов на самостоятельную работу обучающихся. Форма промежуточной аттестации – зачет в 2-м семестре; экзамен в 3-м семестре.

Вид учебной работы	Всего, ак. ч	Семестр 2	Семестр 3
Контактная работа, всего	**120**	**64**	**56**
в т.ч. занятия лекционного типа	60	32	28
в т.ч. лабораторные занятия	0	0	0
в т.ч. практические занятия	60	32	28
в т.ч. семинары	0	0	0
Самостоятельная работа обучающегося	24	8	16
Форма промежуточной аттестации	**–**	**зачет**	**экзамен**
Общая трудоемкость, ак. ч / з.е.	144 / 4	72 / 2	72 / 2

Часы промежуточной аттестации отдельно не выделяются; дополнительная нагрузка сверх общей трудоемкости дисциплины не вводится.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

5.1. Структура дисциплины (модуля) по темам (разделам)

Тема / раздел	Лек.	Лаб.	Практ.	Сем.	СРС	Всего	Контроль
Семестр 2							
Тема 1. Введение: вычислительные методы в drug discovery	4	0	4	0	1	9	практическое задание
Тема 2. Базы данных белков, лигандов и активности	5	0	5	0	1	11	практическое задание
Тема 3. Структура белков-мишеней: моделирование и анализ	6	0	6	0	2	14	практическое задание
Тема 4. Молекулярный докинг	6	0	6	0	2	14	практическое задание
Тема 5. Хемоинформатика: дескрипторы, сходство, химическое пространство	6	0	6	0	1	13	практическое задание
Тема 6. Виртуальный скрининг	5	0	5	0	1	11	практическое задание
Итого, семестр 2	**32**	**0**	**32**	**0**	**8**	**72**	**зачет**
Семестр 3							
Тема 7. Молекулярная динамика и оценка свободной энергии	7	0	7	0	4	18	практическое задание
Тема 8. QSAR и прогнозирование ADMET	6	0	6	0	4	16	практическое задание
Тема 9. Машинное и глубокое обучение в дизайне лекарств	7	0	7	0	4	18	практическое задание
Тема 10. Проект: in silico поиск ингибиторов заданной мишени	8	0	8	0	4	20	проект, презентация; комплексный кейс
Итого, семестр 3	**28**	**0**	**28**	**0**	**16**	**72**	**экзамен**
Итого	**60**	**0**	**60**	**0**	**24**	**144**	

5.2. Содержание разделов (тем) дисциплины

№ п/п	Наименование темы	Содержание темы
1	Тема 1. Введение: вычислительные методы в drug discovery	Место in silico методов в конвейере разработки ЛС. Успешные примеры. Рабочая среда: Linux, Python, Jupyter, RDKit.
2	Тема 2. Базы данных белков, лигандов и активности	UniProt, PDB, AlphaFold DB, ChEMBL, PubChem, BindingDB, DrugBank. Форматы SMILES, SDF, MOL2, PDB/mmCIF.
3	Тема 3. Структура белков-мишеней: моделирование и анализ	Гомологическое моделирование, AlphaFold2/3 и родственные методы, оценка качества модели. Поиск и характеристика сайтов связывания (fpocket).
4	Тема 4. Молекулярный докинг	Алгоритмы поиска и функции оценки. AutoDock Vina, Glide-подобные подходы. Подготовка белка и лиганда. Редокинг, валидация протокола. Ограничения докинга.
5	Тема 5. Хемоинформатика: дескрипторы, сходство, химическое пространство	Молекулярные дескрипторы и отпечатки (ECFP), коэффициент Танимото, кластеризация, визуализация химического пространства (PCA, t-SNE, UMAP).
6	Тема 6. Виртуальный скрининг	Лиганд-ориентированный (фармакофор, сходство) и структурно-ориентированный скрининг. Библиотеки (ZINC, Enamine REAL). Оценка обогащения (ROC AUC, EF).
7	Тема 7. Молекулярная динамика и оценка свободной энергии	Силовые поля, GROMACS/OpenMM. Анализ траекторий. MM/PBSA, FEP – принципы и применение.

№ п/п	Наименование темы	Содержание темы
8	Тема 8. QSAR и прогнозирование ADMET	Построение и валидация QSAR-моделей (OECD-принципы), область применимости. Прогнозирование растворимости, проницаемости, hERG, токсичности.
9	Тема 9. Машинное и глубокое обучение в дизайне лекарств	Классические модели (RF, GBM), графовые нейросети, языковые модели для молекул и белков. Генеративные модели de novo дизайна. Активное обучение.
10	Тема 10. Проект: in silico поиск ингибиторов заданной мишени	Выполнение мини-проекта: выбор мишени, подготовка данных, виртуальный скрининг, отбор и обоснование кандидатов, отчет и презентация. Комплексное задание: Для новой мишени предложены 50 молекул по результатам виртуального скрининга. Сформируйте план их экспериментального подтверждения и правила интерпретации отрицательных результатов.

6. Фонд оценочных средств (ФОС, оценочные и методические материалы) для оценивания результатов обучения по дисциплине (модулю)

6.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости

Текущий контроль успеваемости осуществляется в формах: практические задания в компьютерном классе, тестирование, мини-проект с защитой. Полный комплект оценочных материалов с ключами и эталонами ответов приведен в Фонде оценочных средств по дисциплине (отдельный документ). Примеры заданий:

1. Формат линейной записи химической структуры:
а) VCF; б) PDB; в) FASTA; г) SMILES

2. Коэффициент Танимото используется для оценки:
а) качества модели белка; б) растворимости; в) сходства молекул по отпечаткам; г) энергии связывания

3. Редокинг (redocking) применяют для:
а) синтеза лиганда; б) расчета pKa; в) определения растворимости; г) проверки способности протокола воспроизвести известную позу лиганда

Пример ситуационной задачи: Для мишени известна кристаллическая структура с лигандом и 300 активных соединений в ChEMBL. Требуется отобрать 50 соединений для закупки из библиотеки 5 млн соединений. Предложите стратегию.

Примерные темы докладов (рефератов, эссе, проектов):

1. AlphaFold и его влияние на структурный дизайн ЛС
2. Сверхбольшие библиотеки и виртуальный скрининг
3. Генеративные модели в de novo дизайне
4. FEP в оптимизации соединений-лидеров

6.2. Оценочные материалы для проведения промежуточной аттестации

Промежуточная аттестация в 2-м семестре проводится в форме: зачет (по результатам текущего контроля и собеседования по вопросам к зачету).

Промежуточная аттестация в 3-м семестре проводится в форме: экзамен (устно по билетам, билет содержит два теоретических вопроса и ситуационную задачу).

Примерный перечень вопросов к промежуточной аттестации:

1. Роль вычислительных методов в поиске ЛС.
2. Базы данных структур белков и лигандов.
3. Форматы представления молекул.
4. Методы предсказания структуры белков.
5. Поиск сайтов связывания.
6. Молекулярный докинг: алгоритмы и функции оценки.
7. Валидация протокола докинга.
8. Молекулярные дескрипторы и отпечатки.
9. Меры сходства и химическое пространство.
10. Лиганд-ориентированный виртуальный скрининг.
11. Структурно-ориентированный виртуальный скрининг.
12. Молекулярная динамика: силовые поля и анализ траекторий.
13. Методы оценки свободной энергии связывания.
14. QSAR: построение и валидация моделей.
15. Прогнозирование ADMET in silico.
16. Машинное обучение в хемоинформатике.
17. Генеративные модели для дизайна молекул.
18. Воспроизводимость вычислительных исследований.
19. Обоснуйте план решения комплексной задачи по МПК из раздела 4.3 ФОС и ограничения предлагаемых выводов.

6.3. Шкала и критерии оценивания результатов обучения

Для оценивания результатов обучения по дисциплине используется шкала, соотнесенная с индикаторами достижения компетенций. При промежуточной аттестации в форме зачета оценка «зачтено» соответствует уровням «удовлетворительно», «хорошо», «отлично», оценка «не зачтено» – уровню «неудовлетворительно».

Результат обучения	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
Знания	Отсутствие знаний или фрагментарные знания	Общие, но не структурированные знания	Сформированные, но содержащие отдельные пробелы знания	Сформированные систематические знания
Умения	Отсутствие умений или частично освоенное умение	В целом успешное, но не систематически осуществляемое умение	В целом успешное, но содержащее отдельные пробелы умение	Успешное и систематическое умение
Навыки (владение, опыт деятельности)	Отсутствие навыков или фрагментарное применение	В целом успешное, но не систематическое применение	В целом успешное, но сопровождающееся отдельными ошибками применение	Успешное и систематическое применение

7. Ресурсное обеспечение

7.1. Перечень основной и дополнительной литературы

Основная литература:

1. Лутц,Марк. Изучаем Python : Т. 1 / Марк Лутц ; пер. с англ. и ред. Ю. Н. Артеменко - 5-е изд.. М. СПб. : Диалектика, 2020. URL: https://nbmgu.ru/order/storing.aspx?uid=350a0e74-35a3-463d-9a16-f191e093b1c7&cat=BOOK&db=BOOK. НБ МГУ: Шифр: 5ВГ66, Л-865.

Дополнительная литература:

1. Головин А.В. Поиск новых биоактивных молекул и хемоинформатика. Молекулярное моделирование в применении к биомолекулам. ФББ МГУ, 2024 [Электронный ресурс]. URL: https://kodomo.fbb.msu.ru/~golovin/mmlectures/ll2.pdf.
2. Python Software Foundation. The Python Tutorial [Электронный ресурс]. URL: https://docs.python.org/3/tutorial/.

Библиотечные издания используются в соответствии с правилами доступа НБ МГУ. Электронные нормативные документы и открытая учебная документация выделены отдельно; сведения об источниках и условиях доступа включены в реестр ресурсного обеспечения.

7.2. Перечень профессиональных баз данных и информационных справочных систем

– Головин А.В. Поиск новых биоактивных молекул и хемоинформатика. Молекулярное моделирование в применении к биомолекулам. ФББ МГУ, 2025 [Электронный ресурс]. URL: https://kodomo.fbb.msu.ru/~golovin/mmlectures/ll2.pdf

– Python Software Foundation. The Python Tutorial [Электронный ресурс]. URL: https://docs.python.org/3/tutorial/

– Каталог Научной библиотеки МГУ: https://nbmgu.ru/search/

7.3. Перечень лицензионного и (или) свободно распространяемого программного обеспечения

Linux, Python (Jupyter, RDKit, scikit-learn, PyTorch), AutoDock Vina, Open Babel, PyMOL (open-source), GROMACS, fpocket, KNIME (свободное ПО).

7.4. Описание материально-технической базы

Требования к условиям реализации дисциплины:

Компьютерный класс с рабочими станциями на ОС Linux, мультимедийный проектор. Вычислительные ресурсы с GPU предусматриваются для заданий, требующих аппаратного ускорения; конкретная конфигурация и доступ фиксируются в перечне оборудования.

8. Язык преподавания

Русский.

9. Преподаватель (преподаватели)

Головин Андрей Викторович, доктор химических наук, профессор

10. Разработчик (разработчики) программы

ФИО, ученая степень, должность, подразделение:
Головин Андрей Викторович, доктор химических наук, профессор

Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.

Методические рекомендации и условия для лиц с ОВЗ

Для обучающихся с ограниченными возможностями здоровья и инвалидов формы проведения текущего контроля и промежуточной аттестации устанавливаются с учетом индивидуальных психофизических особенностей (устно, письменно, с использованием электронной информационно-образовательной среды); при необходимости предоставляется дополнительное время для подготовки ответа.